Asad, Y., Ahmad, S., Rungrotmongkol, T., Ranaghan, K. E., & Azam, S. S. (2018). Immuno-informatics driven proteome-wide investigation revealed novel peptide-based vaccine targets against emerging multiple drug resistant Providencia stuartii. *Journal of Molecular Graphics and Modelling*, *80*, 238-250. https://doi.org/10.1016/j.jmgm.2018.01.010

Title Page

Manuscript title: Immuno-informatics Driven Proteome-wide Investigation Revealed Novel Peptide-based Vaccine Targets Against Emerging Multiple Drug Resistant *Providencia stuartii*

Authors: Yelda Asad[a,1], Sajjad Ahmad[a,1], Thanyada Rungrotmongkol[b,c], Kara E. Ranaghan[d], Syed Sikander Azam[a,b,✉]

[1] These authors contributed equally to this work.

Affiliation:

[a]Computational Biology Lab, National Center for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan.

[b]Structural and Computational Biology Research Group, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand.

[c]Ph.D. Program in Bioinformatics and Computational Biology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand.

[d]Centre for Computational Chemistry, University of Bristol, Bristol, United Kingdom.

Corresponding Author:

✉Syed Sikander Azam

E-mail: ssazam@qau.edu.pk; syedazam2008@gmail.com

Phone: 0092-51-906 44130

Computational Biology Lab, National Center for Bioinformatics, Quaid-i-Azam University, Islamabad 45320, Pakistan

Abstract

The bacterium *Providencia stuartii*, is associated with urinary tract infections and is the most common cause of purple urine bag syndrome. The increasing multi-drug resistance pattern shown by the pathogen and lack of licensed vaccines make treatment of infections caused by *P. stuartii* challenging. As vaccinology data against the pathogen is scarce, an *in silico* proteome based Reverse Vaccinology (RV) protocol, in combination with subtractive proteomics is introduced in this work to screen potential vaccine candidates against *P. stuartii*. The analysis identified three potential vaccine candidates for designing broad-spectrum and strain-specific peptide vaccines: FimD4, FimD6, and FimD8. These proteins are essential for pathogen survival, localized in the outer membrane, virulent, and antigenic in nature. Immunoproteomic tools mapped surface exposed and non-allergenic 9mer B-cell derived T-cell antigenic epitopes for the proteins. The epitopes also show stable and rich interactions with the most predominant HLA allele (DRB1*0101) in the human population. Metabolic pathway annotation of the proteins indicated that fimbrial biogenesis outer membrane usher protein (FimD6) is the most suitable candidate for vaccine design, due to its involvement in several significant pathways. These pathways include: the bacterial secretion system, two-component system, β-lactam resistance, and cationic antimicrobial peptide pathways. The predicted epitopes may provide a basis for designing a peptide-based vaccine against *P. stuartii*.

Keywords: *Providencia stuartii*; Reverse vaccinology; Subtractive proteomics; Epitopes; Fimbrial biogenesis outer membrane usher protein.

1. Introduction

Providencia stuartii is an emerging bacterial pathogen and is the most resistant of all *Providencia* species [1]. The pathogen is an etiological agent of complicated urinary tract infections [2], bacteremia [3], diarrhea [4], meningitis [5], wound infections [6], peritonitis [7], endocardium [8], conjunctivitis [9], renal abscesses [10], and septicemia [11]. In addition, *P. stuartii* is the most common cause of purple urine bag syndrome (PUBS), where purple discoloration of urine occurs [12]. The pathogen is a urease producer that significantly increases the incidence of urolithiasis [13]. Antibiotic therapy against *P. stuartii* is losing its efficacy as the majority of strains are resistant to several antimicrobial agents, such as: aminopenicillins, first-generation cephalosporins, fosfomycin, imipenem, polymyxin, penicillin, sulphamethoxazole, tigecycline, and tetracycline [14-17]. *P. stuartii* is naturally resistant to colistin and tigecycline, which are considered as the last-resort treatment for carbapenem-resistant Enterobacteriaceae (CRE) infections. In other words, there is no appropriate antibiotic available for treatment of carbapenem-resistant *P. stuartii* [18, 19]. Moreover, in recent years, the emergence of extended-spectrum β-lactamase (ESBL)-producing strains have made this bacterium a focus of great interest as researchers to attempt find novel routes to tackle this pathogen [14, 15].

Designing vaccines through a classical vaccinology approach is a challenging task. There are several limitations associated with this method, including: culturing problems, long duration, inaccurate and variable products, insufficient attenuation, expensive procedures, less immunogenicity, and hypersensitivity of the antigens [20]. The availability of genomic information allows the identification of suitable vaccine proteins *in silico* without the need to cultivate the pathogen. A proteome based approach for screening virulent antigens, Reverse Vaccinology (RV), has received more attention in recent years and has been used for the identification of vaccine proteins against different pathogens [21, 22, 23, 24, 25]. The RV approach was first applied to the bacterial pathogen Meningococcus B (MenB)[26], where RV played a significant role in screening for an antigen with the broadest bactericidal activity and ultimately resolved the long journey of MenB vaccine development [25]; [27]. RV has also been applied to many other bacterial pathogens, including group A Streptococcus [28], antibiotic-resistant *Staphylococcus aureus* [27], *Streptococcus pneumonia* [27], and Chlamydia [27]; [29]. The efficacy of peptide or subunit based vaccines initially identified through a RV protocol has also

been proven experimentally [30]. Very recently, a RV approach was used to screen possible vaccine proteins against *Acinetobacter baumannii,* identifying outer membrane pilus assembly protein (FliF), as a strong candidate for vaccine development. In a murine pneumonia model, the protein FliF was observed experimentally to provide protection of 50 % against lethal doses of *Acinetobacter baumannii* [30]. Parameters that are important in the screening of vaccine candidates include: outer membrane localization, high virulence, few tran-membrane helices, the presence of signal peptides, antigenicity, and conservation [21, 22]; [31]. In the current approach, a RV framework of *in silico* filters combined with subtractive proteomics was utilized to screen putative vaccine candidates in the proteome of *P. stuartii* strain MRSN 2154. The complete genome of *P. stuartii* reference strain MRSN 2154 was isolated from an Afghan burn-patient and sequenced in 2012. The isolate harbors the New Delhi Metallo-β-lactamase-1 gene, which indicated the presence of numerous other antibiotic resistance loci on the plasmid. It was revealed that *P. stuartii* MRSN 2154 has a circular genome of 4,402,109 nucleotides with G/C content of 42%, 4,194 genes, including 4,099 that code for protein, 75 tRNA genes, and 7 rrn operons [32]. The circular genome of *P. stuartii* genome can be seen in **Fig. 1**. The organism proteome was subtracted to screen out non-paralogs, human-non-homologs, and essential proteins of the pathogen. The proteins from subtractive proteomics were mapped for their association with unique pathways of the bacteria. Unique pathways proteins were characterized physicochemically for the identification of proteins considered feasible for wet lab analysis. The availability of 3D crystal structures for the filtered proteins was checked and those with no available structure also underwent a structure prediction and evaluation phase. Protein-protein interactions were explored for their cellular interacting network to understand the impact of protein inhibition on overall survival of the pathogen. In the epitope mapping phase, B-cell derived T-cell peptides for the proteins were predicted [22, 23, 24, 25] and docked into the binding cavity of DRB1*0101 allele to identify their binding mode and interactions [25].

2. Methodology

The complete flow diagram of the research methodology followed in the manuscript is shown in **Fig. 2**.

2.1. Subtractive Proteomics

The process of antigen screening in the pathogen's proteome starts with subtractive proteomics to remove proteins with no relevance to the aim of the study. For this, the complete proteome of *P. stuartii* reference strain MRSN 2154 (UniProt Id: UP000005012) was retrieved from UniProtKB [33], which served as input for subtractive proteomics pipeline.

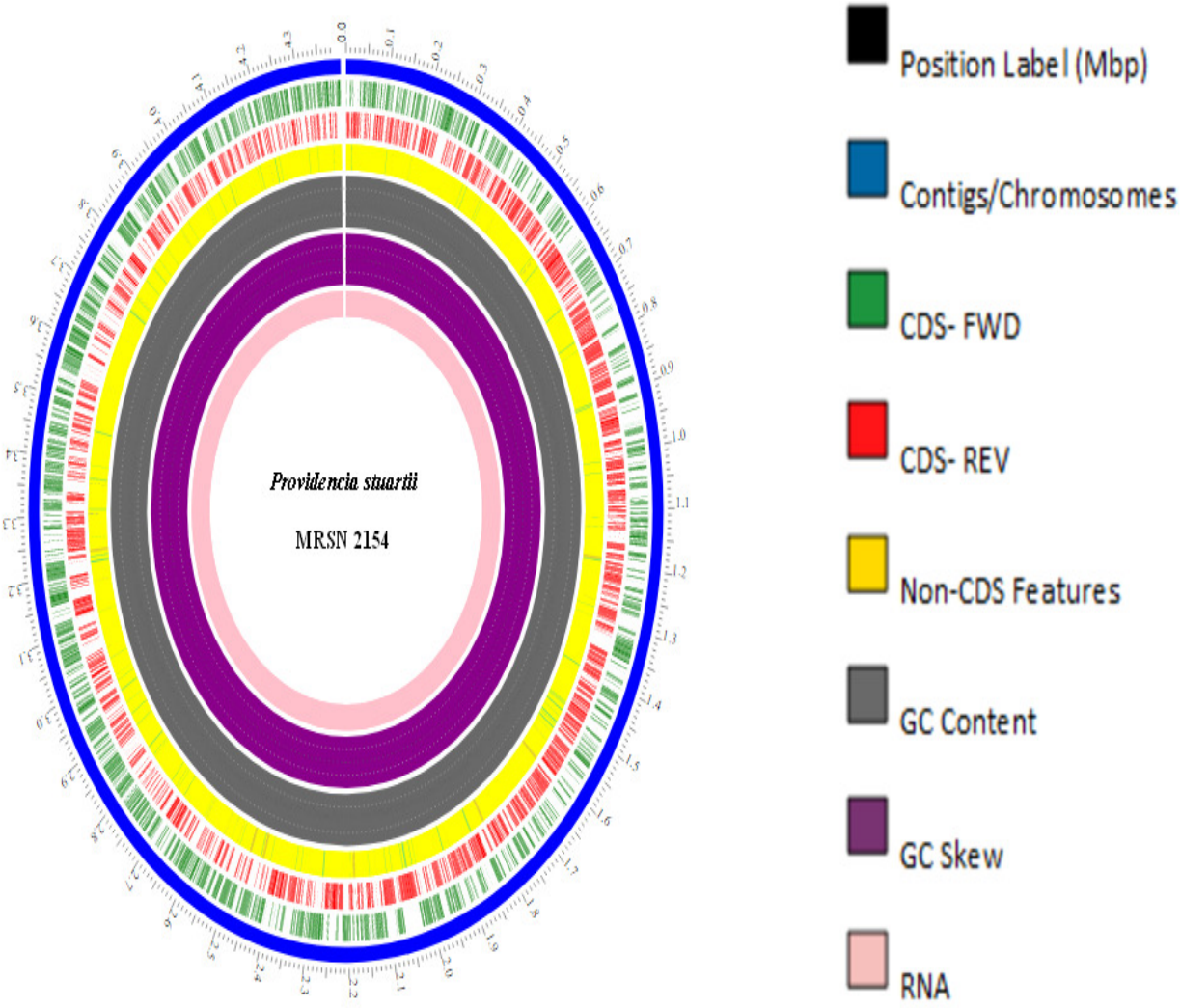


Fig. 1. Circular genome view of *P. stuartii* MRSN 2154.

Initially, paralogous sequences were removed from the genome using the CD-HIT suite (http://weizhongli-lab.org/cd-hit/) [34]. Paralogous proteins are redundant sequences and arise because of duplication events during the evolution process [35].These sequences are less conserved among bacterial species and strains, therefore, not considered as attractive targets for vaccine design [36]. In the CD-HIT assay, the identity cut-off was set to 0.8 (80%), while the remaining parameters were treated as default. The non-paralogous (orthologous) sequences were brought forword to check homology, where sequences were aligned against the human proteome using protein Basic Local Alignment Search tool (BLASTp) [37] available at the National Center for Biotechnology Information (NCBI). The E-value cut-off was set to 10^{-4}, while for scoring and alignment, default parameters were employed. Sequences for which no hit was observed or showed < 35% identity with the human proteome were filtered out as human non-homologous proteins. Host homologous proteins were discarded as they generate autoimmune responses [22]. DEG (database of Essential Genes) (http://www.essentialgene.org/) [38] is a database of genes and proteins fundamental to organism survival. DEG was used for mining essential proteins from the set of non-homologous proteins. A BLASTp search was performed with an E-value cut-off of 10^{-3}, bit score of 100, and sequence identity ≥ 30% against the database to filter proteins fulfilling the threshold values. Essential proteins were targeted as they play a significant role in pathogen's survival and any lethal mutation in these can result in the arrest of cell growth [22]. Surface-exposed (exo-proteome) or secreted proteins (secretome) are suitable vaccine targets due to their frequent contact with the extracellular environment [23]; [39]. In addition, they serve as virulence factors and increase the pathogen's ability to adhere, invade, survive, and proliferate within the host cells [22]; [25]. Therefore, a comparative subcellular localization was performed through three online tools; PSORTb (http://db.psort.org/) [40], CELLO (http://cello.life.nctu.edu.tw/) [41], and CELLO2GO (http://cello.life.nctu.edu.tw/cello2go/) [42].

2.2. Identification of Virulent Proteins

Virulent proteins mediate severe infectious pathways in the host and could be attractive targets for vaccine design [22, 23, 24, 25, 25]; [43]. Several databases are available to predict the virulent potential of a protein. Two databases were used in the current study: the Virulence Factor Database (VFdb) (http://www.mgc.ac.cn/VFs/) [44] and the Microbial Virulence Database (MvirDB)

(http://mvirdb.llnl.gov/) [45]. A BLASTp search was performed against both the databases to remove proteins with an identity ≥ 35% and bit score >100%.

2.3. Metabolic Pathways Annotation

Metabolic pathways annotation was performed for the proteins identified as virulent, to shed light on pathogen pathways crucial for their survival and infectivity [35]. Metabolic pathway analysis was performed using KAAS (KEGG Automatic Annotation Server) (http://www.genome.jp/tools/kaas/) [46], maintained by the KEGG (Kyoto Encyclopedia of Genes and Genomes) database [47].

2.4. Cellular Interactome Analysis

Unveiling the impact of protein-inhibition on the survival of an organism is vital when prioritizing vaccine proteins [22]; [48]. The STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (http://string-db.org/) database was used to disclose the interaction network of the proteins identified by previous screening stages, at the cellular level [49]. The STRING database aims to provide a critical assessment and integration of protein-protein interactions, including both direct (physical) and indirect (functional) associations. Only interactions having an interaction score of 0.4 were considered.

2.5. Vaccine Protein Prioritization

The virulent proteins identified in the previous screening phase were evaluated for antigenicity using VaxiJen (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) [50]**.** The antigenic potential of proteins is imperative for stimulating the human immune system. Proteins having VaxiJen score ≥ 0.4 were considered antigenic and prioritized further. Physicochemical characterization of proteins at the design and development stage is of utmost importance as it aids subsequent wet lab studies [22, 23, 24, 25]. This was based on several criteria including: molecular weight, number of transmembrane helices, and adhesion probability. The molecular weight of the proteins was calculated through ProtParam [51], while numbers of transmembrane helices were evaluated using the TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) [52] and HMMTOP (http://www.enzim.hu/hmmtop/) [53] tools. Proteins having molecular weight < 110 kDa are considered as good targets as such proteins are easy to purify during experimental procedures [22,

23, 24, 25]. Knowledge of the numbers of trans-membrane helices is important, as proteins with a maximum of 2 are more feasible for cloning and expression analysis [22, 23, 24, 25]. Adhesion probability was finally explored using the Vaxign (http://www.violinet.org/vaxign/) server [54]. Adhesion proteins aid in bacterial adherence to host tissues and subsequent colonization [55].

2.6. B-cell derived T-cell Epitope Mapping

In epitope mapping, B-cell derived T-cell epitopes for the proteins were predicted. BCPred (http://ailab.ist.psu.edu/bcpred/predict.html) [56] was used to predict 20-mer B-cell epitopes with a threshold score > 0.8. The antigenicity of each epitope was further assessed using VaxiJen [50] and those having antigenic value ≥ 0.4 were analyzed for surface exposure using TMHMM [52]. For effective vaccine design, the prediction of epitopes that bind to both classes of MHC is imperative [22, 23, 24, 25]. Binding alleles of the epitopes were determined through Propred1 (http://www.imtech.res.in/raghava/propred1/) [57] and Propred (http://www.imtech.res.in/raghava/propred/) [58]. T-cell epitopes, which bind to more than 15 MHC alleles, particularly DRB1*0101, were considered [59]. The half maximal inhibition concentration (IC_{50}) score of the mapped T-cell epitopes was calculated using MHCPred [60] and those with a score < 100 nM were further analyzed for virulence using VirulentPred [61] with a cut-off ≥ 0.5.

2.7. Protein and Epitope Allergenicity

It is important to incorporate an allergenicity check in the design framework as an increased number of vaccines are now reported to cause allergic reactions [62]. Allerdictor (http://allerdictor.vbi.vt.edu/) [63] was used to predict the allergenicity of the proteins and their respective epitopes.

2.8. Epitope Conservation and Evolutionary Studies

In order to design a broad-spectrum and strain-specific vaccine, the predicted epitopes were inspected for conservation among completely sequenced strains using CLC Sequence Viewer (www.clcbio.com) [64]. To understand the evolution pattern of *P. startuii* strains, a 16SrRNA-based evolutionary tree was constructed using MEGA6 [65]. Since a 16SrRNA sequence is subject to fewer variations, they are often targeted for phylogenetic studies [23].

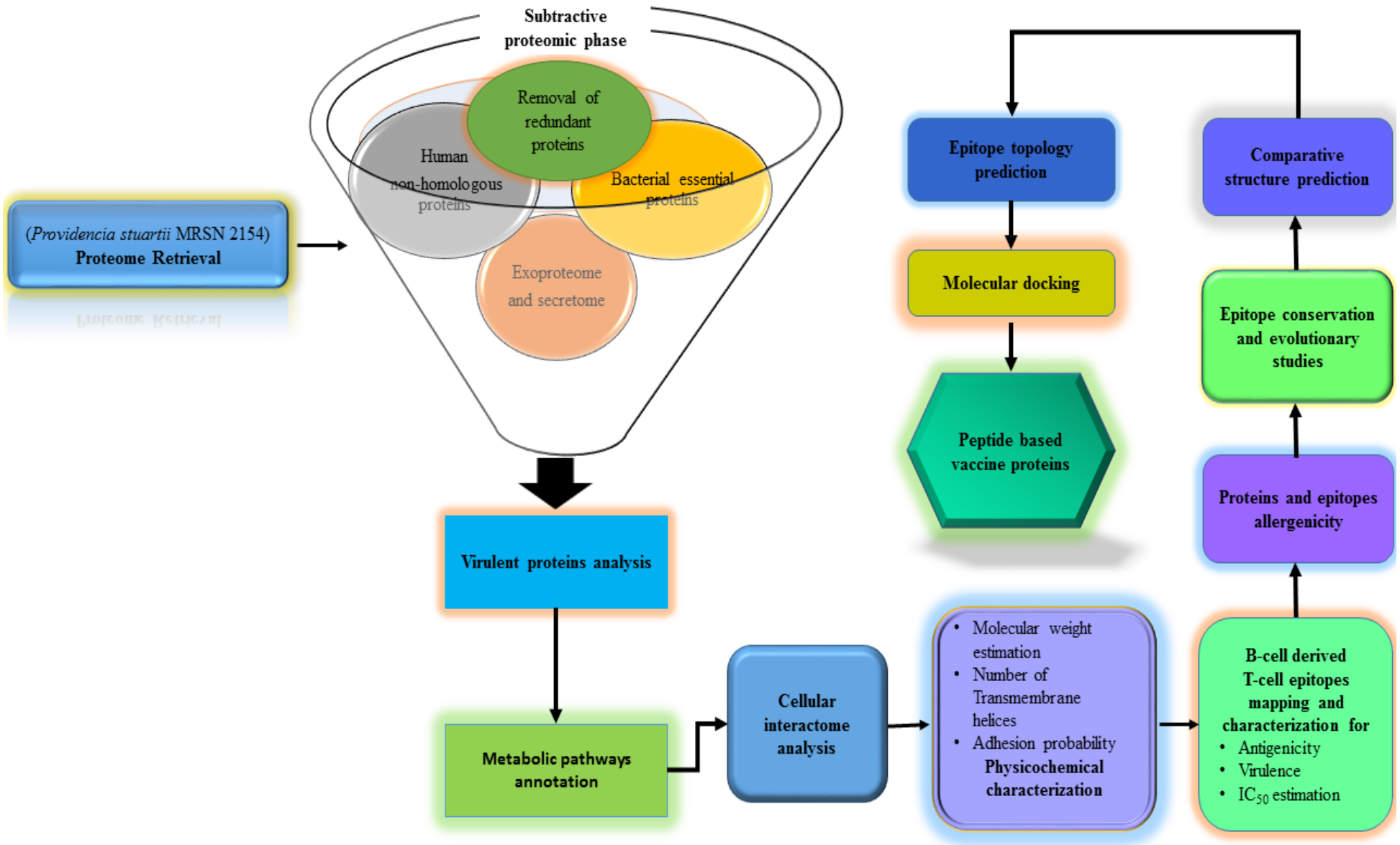


Fig. 2. The adopted methodology outlining major activities performed during different phases.

RNAmmer (http://www.cbs.dtu.dk/services/RNAmmer/) [66] was used with default settings for prediction of 16SrRNA sequence and those with the score of higher than 1700 were subjected to multiple sequence alignment and phylogenetic tree construction [23]. The multiple sequence alignment was performed using ClustalW and a phylogenetic tree was constructed based on the Neighbor-Joining method [67].

2.9. Protein Structure Prediction

In order to visualize the topology of predicted epitopes on the protein surface, the structure of the proteins was predicted [22, 23, 24, 25]. Proteins with no structural information available in protein data bank (PDB) [68] were predicted through a comparative structure prediction approach using the following online tools: Phyre2 [69], SwissModel [70], Intfold2 [71], Modweb [72], and I-Tasser [72]. Each of the predicted structures then underwent structure evaluation analysis through the following servers: PROCHECK [74], ProSA [75], ERRAT [76], and Verify3D [77]. The structure with the fewest residues in the disallowed region of a Ramachandran plot and having most of its residues mapped, was selected and its energy optimized using UCSF Chimera [78]. Structure minimization was performed for 1500 steps (750 steepest descent and 750 conjugate gradient with a step size of 0.02 Å) assigning Gasteiger charges under the Tripos Force Field (TFF) [25].

2.10. Epitope Topology Prediction

The topology of predicted epitopes was visualized using the Pepitope server (http://pepitope.tau.ac.il/) [79]. This analysis was vital to gain insights about the 3D structure of the protein and the epitope's exomembrane topology [22, 23, 24, 25].

2.11. Molecular Docking

Docking of epitopes with the most common binding allele, DRB1*0101, was performed using Autodock Vina [80] and GalaxyPepDock server [81]. The 3D structure of the allele was retrieved from the PDB (PDB ID "1AQD") and docked with the prioritized epitopes. The peptide-protein complexes were ranked based on the predicted binding affinity of the peptide in the protein active site in Autodock Vina. Results from GalaxyPepDock were assessed based on similarity score,

estimated accuracy, and interaction score. The complexes were analyzed for binding mode using UCSF Chimera [78], and binding interactions using LigPlot [82].

2.12. Validation through a Positive and Negative Control

The functionality of the framework and validation of the screened vaccine proteins was achieved using a positive and negative control. The same pipeline was applied to *Helicobacter pylori* strain 26695 (positive control) [22, 23] and *Mycoplasma pneumoniae* strain ATCC 15531 (negative control) for predicting vaccine proteins against them.

3. Results

3.1. Proteome Subtraction

RV in combination with subtractive proteomics is a hybrid approach that can be explored for screening of potential peptide-based vaccine candidates in the proteome of bacterial pathogens [21]. Using this comprehensive approach, we screened antigenic epitopes in the complete proteome of *P. stuartii* reference strain MRSN 2154. First, the proteome was filtered through CD-Hit server, which removes paralogous sequences from the proteome. It was revealed by the CD-Hit server that out of total 4220 proteins, 4163 proteins were non-paralogous. Paralogous sequences are present in the proteome because of duplication events and do not form the core proteome. Therefore, it was considered logical to remove such sequences due to their insignificance in the context of vaccine design. To avoid autoimmune reactions, human homologs were removed from the non-paralogous sequences by performing a BLASTp search against the human proteome [23]; [37]. This analysis identified 3561 proteins which were subsequently subjected to DEG analysis [38], revealing 1062 essential proteins. The check for essential proteins is a key step in the prediction of target proteins, as essential proteins are important factors for the survival, adhesion, entry into the host, infection, as well as persistence of the pathogen inside the host [22, 23, 24, 25].

The identified essential proteins were also examined for subcellular localization using three online servers: PSORTb [40]**,** CELLO [41], and CELLO2GO [42]. Three servers were used as the PSORTb server could not predict the localization for some of the proteins. According to PSORTb, only 1 % of the proteins were found to be located in extracellular regions, 17% were outer membrane, and 82% were of unknown localization. CELLO indicated as 18% extracellular and 82% as outer-membrane proteins, whilst CELLO2GO predicted 23% to be extracellular and 77% to be outer membrane proteins. By comparing the results of three tools, 25 proteins were predicted as outer membranous and one extracellular. Targeting extracellular and outer membrane proteins is vital, as they can be effective candidates due to their exposure to the extracellular environment and interaction with their surrounding biotic and abiotic components [22, 23, 24, 25]. In addition, their role in pathogen virulence, adherence, host cell invasion, and proliferation make them excellent targets for vaccine design [22, 23, 24, 25]**.**

3.2. Virulent Proteins Identification

Extraction of virulent proteins from 25 extracellular and outer membrane proteins was done using a BLASTp search of VFDB [44] and MvirDB [45] against their core data set. The analysis revealed 13 virulent proteins: FimD1, FimD2, FimD3, FimD4, FimD5, FimD6, FimD7, FimD8, TolC, OmpF1, OmpF2, TC.OOP, and MepM **(Table.1)**.

Table 1. Virulent proteins of *P. stuartii* identified through virulent protein analysis.

UniProt ID	KEGG Description	KEGG Name
A0A140STA8	Outer membrane usher protein	FimD1
A0A140NFQ5	Outer membrane usher protein	FimD2
A0A140NKU8	Outer membrane usher protein	FimD3
A0A140STH0	Outer membrane usher protein	FimD4
A0A140NJK0	Outer membrane usher protein	FimD5
A0A140NPZ6	Outer membrane usher protein	FimD6
A0A140NJ67	Outer membrane usher protein	FimD7
A0A140NNX3	Outer membrane usher protein	FimD8
A0A140NQ20	Outer membrane protein	TolC
A0A140NKJ2	Outer membrane pore protein F	OmpF1
A0A140NQ63	Outer membrane pore protein F	OmpF2
A0A140NTJ5	OmpA-OmpF porin, OOP family	TC.OOP
A0A140NV54	Murein DD-endopeptidase	MepM

FimD family proteins (FimD1, FimD2, FimD3, FimD4, FimD5, FimD6, FimD7, and FimD8) are outer membrane molecular usher proteins and play a crucial role in chaperone uncapping and pilus subunits translocation across the outer membrane during pili assembly [83]. TolC has a role in the exclusion of diverse molecules such as protein toxins and antibacterial drugs from the cell [84]. OmpF is a porin protein, which acts as passive diffusion channel for nutrients, small molecules, toxic salts, and antibiotics [85]. MepM functions in peptidoglycan maturation and recycling as

well as in cell separation [86]. TC.OOP functions in the process of epithelial invasion and apoptosis [87].

3.3. Metabolic Pathway Analysis

The 13 proteins identified by previous rounds of screening were mapped to metabolic pathways using KAAS. Out of 13 proteins, only 5 proteins were identified as being associated with several metabolic pathways of the pathogens. These proteins include FimD1, FimD4, FimD6, FUB7, and FimD8. FimD6 is involved in a bacterial secretion system, two-component system, plant-pathogen interaction, beta-lactam resistance, and synthesis of cationic antimicrobial peptides. Bacteria have developed several routes to enter into mammalian host cells, destruct tissue sites, and dampen the immune system response. One such way is the secretion of proteins across phospholipid membranes using protein secretion systems and is vital for the survival of bacterial pathogens. The secreted proteins have many roles in bacterial virulence including attachment, intoxicating, and subsequent disruption of target cells function. Secretion systems also play a significant role in scavenging resources in an environmental niche, thus making the host the fittest in the survival race [88]. Two-component systems facilitate bacteria in signaling events, including cell-cell communication, adaptation to environments, and pathogenesis [89]. This protein is also involved in exhibiting resistance to β-lactams, which include penicillin, cephalosporin, carbapenem, and monobactam [90]. The involvement of FimD6 in several pathways, make it a key target for future vaccine design **(S-Table 1.)**.

3.4. Protein-Protein Interactions (PPIs)

The five proteins identified in the last step were further evaluated for PPIs within the pathogen proteome using STRING [49]. Each protein showed ten interacting partners in their protein network **(S-Table 2.)**. For FimD1, FimD6, and FimD7, the interacting proteins were almost the same, involved in gram-negative bacteria pili assembly, cell wall organization, and chaperone-mediated protein folding [49]; [83]; [91]. Pili or fimbriae is a virulence factor that helps bacteria to colonize host specific tissues and enhance its pathogenesis [49]; [91]. In the case of FimD4, 9 interacting proteins were pili proteins while the remaining one was a cell adhesion protein. For FimD8, 6 proteins were part of cell wall organization, chaperone-mediated protein folding, and pilus organization, while the other four were involved in cell adhesion [49]. Bacteria need monomeric adhesions/invasions or highly complicated macromolecular machines such as type III

secretion systems, and retractile type IV pili to form a complex host-pathogen molecular cross-talk that leads to disruption of cellular functions, and establishment of disease [47].

3.5. Vaccine Protein Prioritization

The five identified proteins were analyzed for antigenicity using VaxiJen. All the five proteins were found to be antigenic; FimD1 (0.5), FimD4 (0.6), FimD6 (0.5), FimD7 (0.6), and FimD8 (0.5). After confirming them as antigenic, the proteins were subjected to transmembrane helices analysis using TMHMM [52] and HMMTOP [53]. It was revealed that FimD4 and FimD6 have one transmembrane helix, whereas candidates FimD1, FimD7 and FimD8have no transmembrane helices. Computing the number of transmembrane helices was important, as proteins having large number of transmembrane helices are difficult to clone and express [22, 23, 24, 25]. As all the proteins have a favorable number of transmembrane helices, they were further analysed for adhesion probability. The values of adhesion probability for FimD1, FimD4, FimD6, FimD7, and FimD8 were 0.5, 0.6, 0.5, 0.6, and 0.5, respectively, indicating their adhesive nature. Adhesive proteins allow bacteria to adhere abiotic surfaces and thus can enhance the pathogenic potential of the bacteria [25]; [54]; [55]. Furthermore, the proteins were physicochemically characterized through Protparam [51]. The Grand average of hydropathicity (GRAVY) index value is a measure of the hydrophobic or hydrophilic nature of a protein. For most proteins, GRAVY values range from -2 to +2, where negative values indicate a hydrophobic nature. The predicted GRAVY indices were in the range -0.2 to -0.4, this indicates that they are hydrophilic and nearly all of their residues are located on the surface, which tends to interact with the host proteins. The instability index of a protein was predicted by calculating the weighted sum of dipeptides that occur more frequently in unstable proteins when compared to stable proteins. An instability index greater than 40 indicates an unstable protein. An instability index of less than 40 was predicted for all the targeted proteins, indicating that they are stable. The theoretical pI value was less than 7 indicating the acidic nature of all the 5 selected proteins. The average aliphatic index value for all the proteins denotes high thermal stability. The molecular weights of the target proteins were calculated. Molecular weight of the targeted proteins in kDa were as follows: FimD1 (92.1), 93.2 (FimD4), 93.3 (FimD6), 78.3 (FimD7), and 96.8 (FimD8). Low molecular weight ($\leq$ 110 kDa) proteins are considered more effective targets, as they are easy to purify and are favored during vaccine development **(Table 2.)**

[22, 23, 24, 25]. The number of prioritized proteins at each step of the pipeline is presented in the form of Venn diagram and can be found in **(Fig. 3.)**.

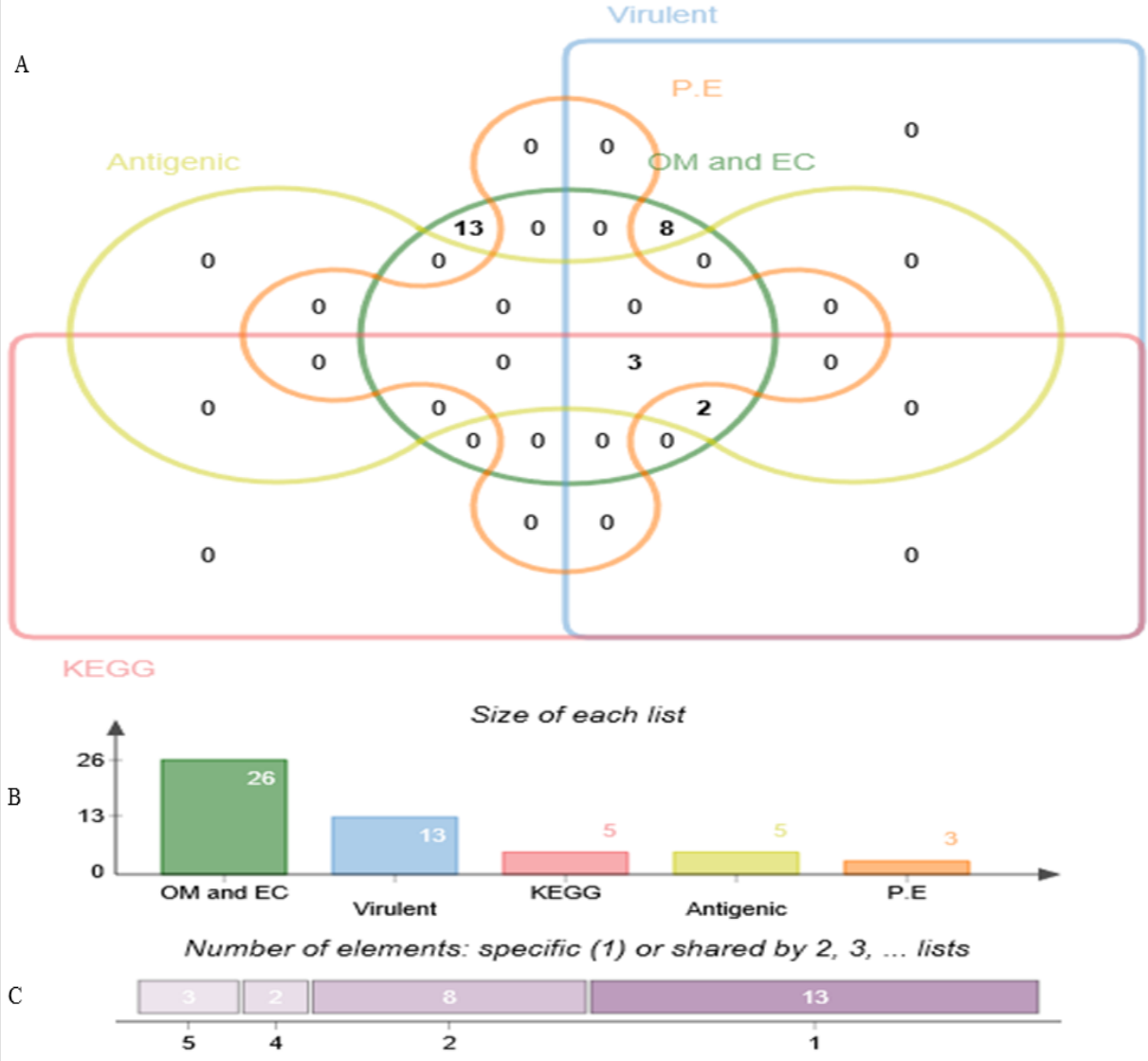


Fig. 3. Classification of the *P. stuartii* proteome. **A.** A representation of the prioritized proteins selected from the proteome of *P. stuartii* on the basis of sub-cellular localization, virulence, KEGG (involvement in metabolic pathways) and antigenic properties. In the Venn diagram, EC (extracellular) and OM (outer membranous) proteins are green , virulent proteins are blue, proteins involved in metabolic pathways (KEGG) are pink, antigenic proteins are yellow and predicted epitopes are orange. The five prioritized proteins share all the properties (outer membrane, virulent, involvement in the metabolic pathway and antigenic). **B.** Shows bar graph for each category of protein **C.** The number of shared proteins and specific proteins in each category.

370 **Table 2.** Prioritized vaccine proteins for antigenic epitope mapping.

371

Protein	Essential	Virulent	VaxiJen Score	TMHMM	HMMTOP	Adhesion Probability	GRAVY	Aliphatic Index	Instability Index	Length	Theoretical pI	Molecular Weight
FimD1	Yes	Yes	0.5	0	0	0.5	-0.4	77.9	25.2	829	6.3	92.1
FimD4	Yes	Yes	0.5	1	1	0.6	-0.2	79.2	35.0	855	6.7	93.2
FimD6	Yes	Yes	0.5	0	1	0.5	-0.3	74.9	29.4	853	6.8	93.3
FimD7	Yes	Yes	0.5	0	0	0.6	-0.4	75.6	32.3	708	6.7	78.3
FimD8	Yes	Yes	0.5	0	1	0.5	-0.4	80.5	26.9	878	5.8	96.8

3.6. Mapping of B-cell Derived T-cell Epitopes

Engineering vaccines that can stimulate both humoral and cell-mediated immunity is imperative in vaccinology [92]. In the current study, we screened epitopes that harbor epitopes of both B-cell and T-cell, signifying their potential of eliciting both arms of immunity i.e. cellular and humoral immunity. We identified first B-cell epitopes and then subjected the B-cell epitopes to T-cell epitope prediction. These B-cell derived T-cell epitopes have the potential to stimulate both types of immunity. Significantly, these epitopes produced specific and selective immune responses and avoid many unwanted allergic and reactogenic responses [22, 23, 24, 25]. The proteins found to have physicochemically favorable properties were, therefore, subjected to a B-cell derived T-cell epitope prediction phase. First, BCPred [56] was used to predict 20-mer B-cell epitopes for each of the targeted proteins. The 5 proteins all have between 18 and 22 epitopes [FimD1=20, FimD4=22, FimD6=18, FimD7=17, FimD8=19].Those epitopes with a BCPred score > 0.8 and a VaxiJen score > 0.4 [51] were selected for exposed topology analysis using TMHMM [43]. Surface exposed epitopes were screened for their binding affinity with MHC-class I and MHC-class II molecules. The Propred1 [57] and Propred [58] servers were used to predict binding of antigenic surface exposed B-cell epitopes to MHC-class I and MHC-class II alleles, especially DRB1*0101 due to its frequent occurrence. Epitopes that bind to the DRB1*0101 allele tend to result in good immunogenic responses. Epitopes that bind to more than 15 MHC class I and II alleles were selected, and evaluated further for IC_{50} values using MHCPRED [60], antigenicity through VaxiJen [50], and virulence via VirulentPred [61]. Those epitopes fulfilling all these criteria were considered. It was found that three epitopes from FimD4 showed binding to more than 15 MHC classes, of which only the best epitope 'FVINDLYPT' was selected due to its binding affinity for 51 MHC alleles (including DRB1*0101), antigenicity value of 0.5, IC_{50} value of 4.2, and a virulence score of 1.0. The epitope determined from FimD6 was 'IQWGNSHSA' which binds to 38 MHC alleles, has a VaxiJen score of 1.2, an IC_{50} value of 54.8 and virulence value of 1.0. Similarly, the 'YGNANVGYS' epitope from FimD8 was considered because of its binding with 54 MHC alleles, antigenicity score of 1.2, IC_{50} value of 19.9, and virulent value of 1.0 **(Table. 3.)**.

403 **Table. 3.** B-cell derived T-cell epitopes from three prioritized vaccine proteins.

404

Protein	B-Cell Epitopes	T-Cell Epitopes	MHC Binding Alleles	VaxiJen Score	VirulentPred	MHCPred IC_{50} Value	Allidictor	Pepitope
FimD4	APGPFVINDLYPTGTAGDLE	FVINDLYPT	51	0.5	1.0	4.2	Non-allergen	All residues exposed
FimD6	PDSGTLLIQWGNSHSASCGA	IQWGNSHSA	38	1.2	1.0	54.8	Non-allergen	All residues exposed
FimD8	YGNANVGYSYTNDLSQMYYG	YGNANVGYS	54	1.2	1.0	19.9	Non-allergen	All residues exposed

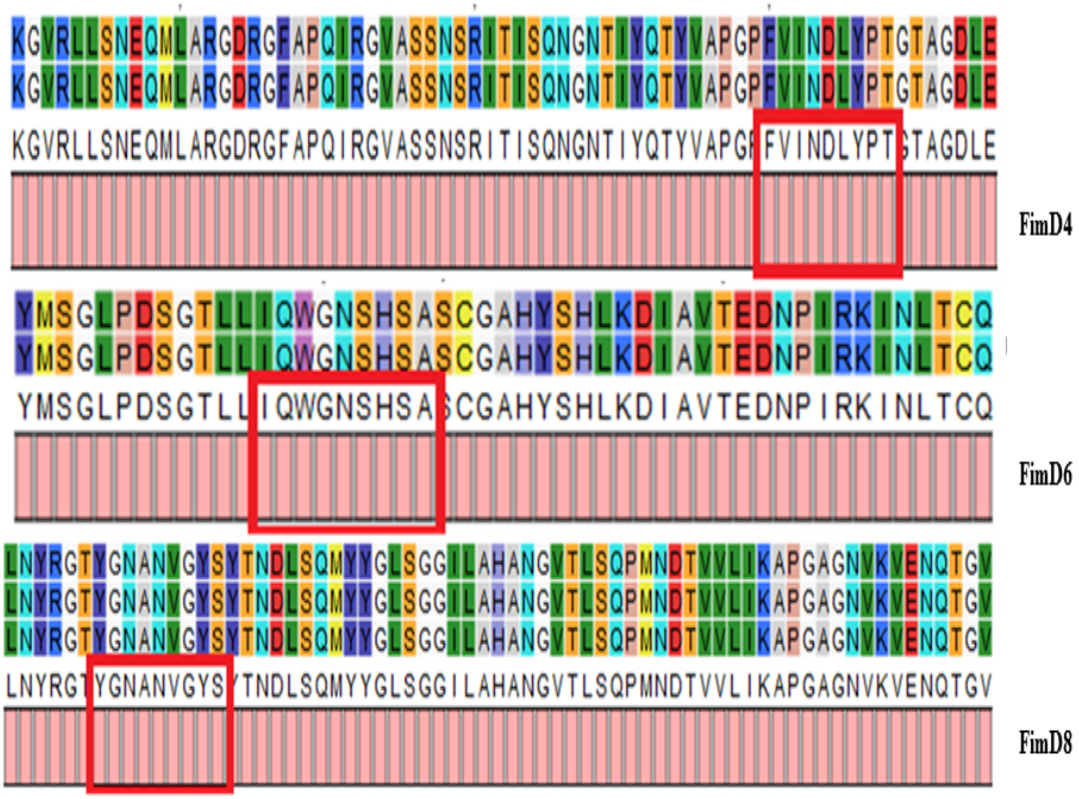
405

3.7. Allergenicity Evaluation

A vaccine based on the whole organism or large proteins are not favored due to induction of non-specific immune responses and the reactogenic and/or allergenic responses associated with them. Such issues can be avoided by designing peptide vaccines based on short peptides capable of inducing highly targeted immune responses with no reactogenic and/or allergenic responses [93]. Through the Allerdictor tool [63], it was found that all the three proteins and their epitopes were non-allergenic. Consequently, these protein epitopes can be considered suitable candidates for peptide-based vaccine design.

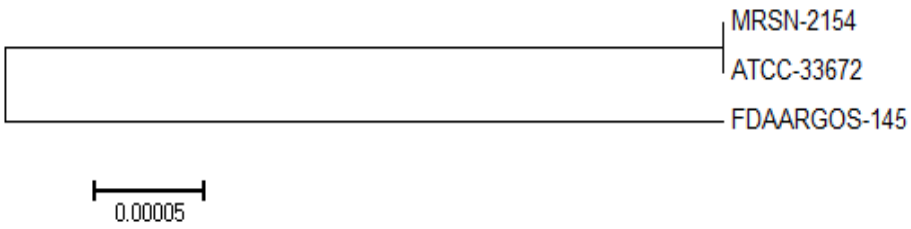
3.8. Conservation and Phylogenetic Analysis

Conservation of epitopes is essential for the design of broad-spectrum vaccines and was thus included in the design framework. The conservation of the target epitopes was analyzed among three fully annotated proteomes of *P. stuartii* using CLC Sequence Viewer. The epitope sequence 'FVINDLYPT' from FimD4 and 'IQWGNSHSA' from FimD6 were conserved in MRSN 2154 and FDAARGOS_145 strains, but was absent in ATCC 33672 and thus, can be targeted for strain-specific vaccines. The epitope sequence 'YGNANVGYS' from FimD8 was predicted to be conserved in all the 3 strains, an attractive target for a broad-spectrum vaccine **(Fig. 4.)**. To understand the evolution pattern, a 16SrRNA-based evolutionary tree was constructed using MEGA6 [65]. The tree illustrated the genetic distances between closely related strains of the pathogen. It was found that *P. stuartii* MRSN 2154 and ATCC-33672 fall in the same clade, showing their phylogenetic closeness, while *P. stuartii* FDAARGOS_145 is expressed as a separate branch of the tree, indicating it is phylogenetically distant from the other two strains **(Fig. 5.)**. The close relationship between *P. stuartii* MRSN 2154 and *P. stuartii* ATCC 33672 can be explained based on the genomic information. It was concluded that both the strains have almost the same genetic contents and can act as model genomes for computational and experimental studies. As described earlier, *P. stuartii* MRSN 2154 has a 4,402,109 circular genome, 43% G/C content and 4,194 predicted genes [32]. Similarly, *P. stuartii* ATCC 33672 has a 4,285,951-bp circular genome, 42% GC content, and 4,094 predicted genes [94].



433

434 **Fig. 4.** Alignment and conservation of selected epitopes from FimD4, FimD6 and FimD8 among
435 three completely annotated strains of *P. stuartii.* The conserved epitopes are highlighted in red
436 boxes.



437

438 **Fig. 5.** The 16SrRNA distance tree representing an evolutionary pattern of *P. stuartii* strains.
439 MEGA6 (molecular evolutionary genetic analysis) was used for multiple sequence alignment of
440 selected 16SrRNA genes through ClustalW and phylogenetic tree construction by the Neighbor-
441 Joining method. The tree has a branch length = 0.00005. The evolutionary distances were
442 calculated using the maximum composite likelihood method and the branch lengths were measured
443 in the number of substitutions per site.

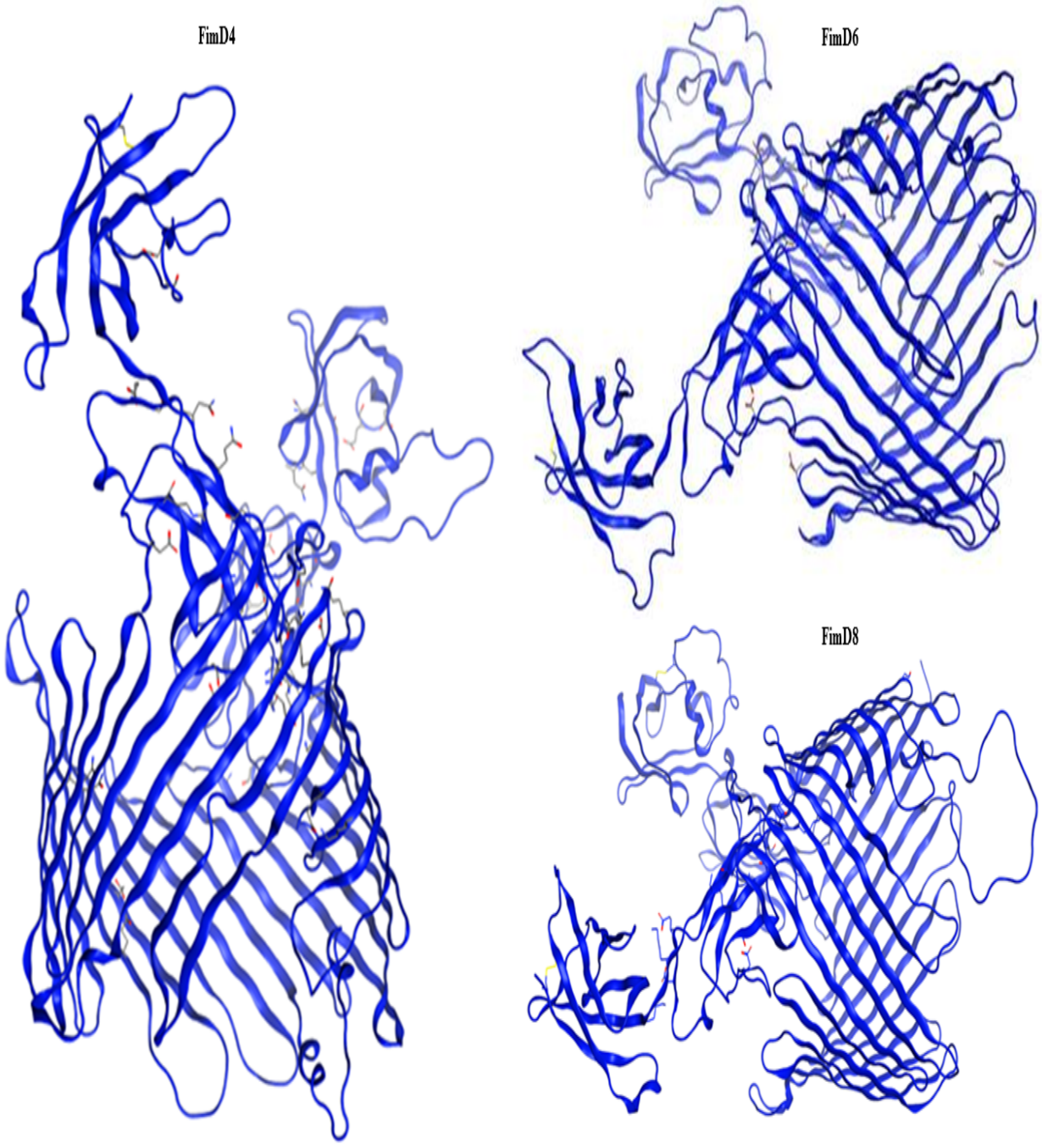
3.9. Prediction and Evaluation of 3D Structures

To view the topology of the predicted epitopes in a subsequent step, the 3D structures of the proteins were needed. The 3D structures of the proteins were predicted using a comparative structure prediction approach. To select the optimum model, the evaluation was done for all the structures predicted using different tools. The most reliable model was selected by analyzing their Ramachandran plot generated by PDBSum [74], the Z-score of ProSA [75], the ERRAT quality score [76], and the Verify3D score [77]. The model generated by Swiss-model [70] was the most reliable for FimD4 with an ERRAT value of 85.1 and a Verify3D value of 85.2. PDBsum scores of the model were as follows: 89.5% residues in most favored regions, 9.5 % residues in additionally allowed regions, 0.60 % residues in generously allowed regions, and 0.4% residues in disallowed regions, G-factor value of -0.1, and Z–score value of -9.2. Statistically, the structure possesses 50.8 % of strand, 2.5 % of alpha helix, 4 % of 3-10 helix, and 42.7 % of others **(S-Fig. 1.)**. The Modweb [72] generated model was selected for FimD6 on the basis of ERRAT value (59.5), Verify3D value (85.7), residues in most favored regions (87.9 %), residues in additionally allowed region (10.6 %), residues in generously allowed regions (1.2 %), residues in disallowed region (0.3 %), G-factor, and Z-score of -0.1 and -9.5, respectively. The secondary structure for FimD6 depicted that like FimD4, it is made up of more strands than helices. Statistically, the protein has 50.1 % of strand, 16 % of helix, 2.1 % of 3-10 helix, and 45 % of others **(S-Fig. 2.)**. Based on the model evaluation results, the model generated by Intfold2 [71] was the best model for FimD8 with ERRAT value (66.8), Verify3D value (80.6), residues in most favored region (91.0 %), residues in additionally allowed regions (7.8 %), residues in generously allowed regions (0.70%), residues in disallowed region (0.5%), G-factor value of -0.1 and Z–score value -8.8. **(S-Table 3)**. The minimized structure of the selected models can be found in **(Fig. 6.)**. The secondary structure of FimD8 also has more B-sheets (47.4 %) compared to alpha helix (1.8 %), 3-10 helix (3.1 %), and others (47.7 %) **(S-Fig. 3.)**.

3.10. Topology of Epitopes

The topology of antigenic and virulent epitopes was evaluated to ensure their surface exposure [22, 23, 24, 25]. This analysis is also assured to validate the non-folding of epitopes inside globular proteins and was performed using Pepitope [79]. This analysis revealed that all the epitopes are surface exposed **(Fig. 7.).** The epitope sequence ‘FVINDLYPT’ of FimD4 was located from

474 position 320 to 328, ‘IQWGNSHSA’ of FimD6 from 817 to 825 and ‘YGNANVGYS’ of FimD8
475 from 677 to 685.



476

477 **Fig. 6.** Minimized structures of the best-predicted structures for prioritized vaccine proteins.

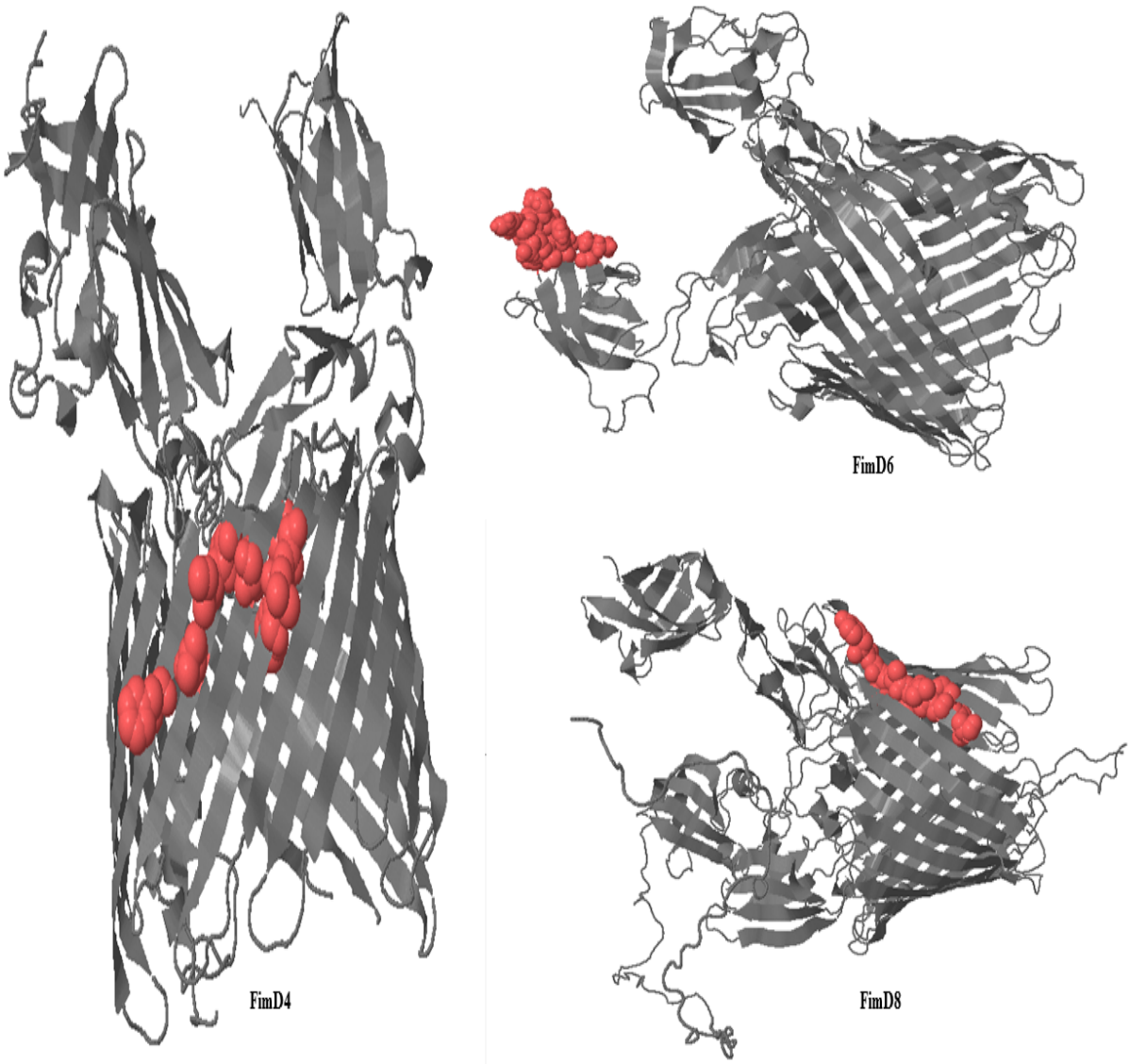


Fig. 7. The topology of predicted antigenic epitopes within the 3D structure of the corresponding proteins. **A)** FimD4 **B)** FimD6 **C)** FimD8.

3.11. Binding Mode and Interaction Analysis

The best peptide-protein complex was selected, and analyzed for binding mode using UCSF Chimera [78], and binding interactions were identified via Ligplot [82]. In general, the peptides were found to bind in the same site, with similar interactions predicted by both Autodock Vina and GalaxyPepDock. A comparative analysis of peptide docking is shown in **Table 4**. For all 3 protein-

peptide complexes the binding affinities and scores from GalaxyPepDock are similar. In general, the complexes predicted by GalaxyPepDock contain more interactions between the protein and the peptide than the structures predicted by Autodock Vina. The Autodock Vina binding affinity of the peptide was -9.4 kcal/mol, while GalaxyPepDock, predicts a similarity score of 0.9, an interaction similarity score of 116.0, and an estimated accuracy of 0.8. In Autodock Vina, 7 hydrogen bonds and 19 hydrophobic interactions were identified, while GalaxyPepDock predicted 8 hydrogen bonds and 21 hydrophobic interactions. Similarly, the FimD6 peptide was found to bind in the same binding pocket as identified in FimD4. The FimD6 peptide formed multiple hydrogen bond and hydrophobic interactions with the active site residues. In Autodock Vina, the peptide formed 7 hydrogen bonds and 13 hydrophobic interactions, while GalaxyPepDock predicted 9 hydrogen bonds and 17 hydrophobic interactions. The Autodock Vina binding energy of the peptide was -7.9 kcal.mol and the GalaxyPepDock scores were 0.9 for protein similarity, 114 for the interaction similarity score, and 0.9 for the estimated accuracy. Lastly, the FimD8 peptide has a binding affinity of -8.8 kcal/mol and similarity score of 0.9, interaction similarity score of 118, and estimated accuracy of 0.9 in GalaxyPepDock. Ligplot analysis revealed 6 hydrogen bonds and 18 hydrophobic interactions for the Autodock Vina structure, while GalaxyPepDock predicted 6 hydrogen bonds and 16 hydrophobic interactions.

Table 4. Comparative analysis of peptide docked complexes interactions in Autodock Vina and GalaxyPepDock.

Receptor	Peptide	Autodock Vina		GalaxyPepDock	
		Hydrogen bonding	**Hydrophobic interactions**	**Hydrogen bonding**	**Hydrophobic interactions**
1AQD	FVINDLYPT (FimD4)	Gln7, Ser51, Tyr58, Trp59, Asn60, Arg69, Arg74	Trp7, Leu9, Phe11, Phe20, Phe22, Glu26, Ile29, Trp41, Phe52, Glu53, Gly56, Trp59, Gln62, Asn67, Ile71, Tyr76, His79, Asn80, Val163	Gln7, Ser51, Asn60, Asn67, Tyr236, Gln 246, Arg247, Asn258	Glu9, Phe22, Phe30, Ala50, Phe52, Glu53, Gly56, Val63, Asp64, Ile70, Leu187, Phe189, Asp233, Gln240, Leu243, Ala250, Thr253, Tyr254, His257, Val261, Phe265
1AQD	IQWGNSHSA (FimD6)	Gln7, Ser51, Glu53, Asn60, Gln68, Thr75, Asn80	Phe11, Phe22, Glu26, Phe32, Trp41, Tyr45, Phe52, Ala57, Gly58, Trp59, Val63, Ala72, Tyr78	Gln7, Arg8, Ser51, Glu53, Asn60, Asp64, Asn67, Arg247, Asn258	Phe22, Ile29, Phe30, Trp41, Ala50, Phe52, Gly56, Ala59, Val63, Leu187, Phe189, Tyr223, Trp237, Tyr254, His257, Val261, Phe265
1AQD	YGNANVGYS (FimD8)	Gln7, Ser51, Asn60, Asn80, Gln68, Arg69	Ile5, Leu9, Phe11, Phe20, Glu26, Cys28, Ile29, Phe30, Trp41, Phe52, Gly56, Trp59, Val63, Asp64, Leu65, Asn67, Trp76, Gly84	Gln7, Asn60, Trp237, Arg247, His257, Asn258.	Glu9, Phe20, Phe22, Ile29, Phe30, Ala50, Phe52, Gly56, Val63, Asn67, Ile70, Phe189, Phe265, Asp233, Tyr236, Gln246

3.12. Validation through a Positive and Negative control

The pipeline employed in the current work was used to analyze the proteome of *H. pylori* for potential vaccine candidates. The pipeline predicted 5 significant potential vaccine proteins against *H. pylori*, which includes VacA, babA,sabA, fecA, and Omp16 **(S-Table 4.)** [22]. Similarly, we used a negative control to further validate our results. The pipeline revealed no potential vaccine proteins against *M. pneumonia*. The CD-hit analysis revealed 606 proteins out of a total of 629, a BLASTp search against the human proteome found only 500 proteins as human non-homologous. Essentiality analysis found 292 proteins as pathogen-essential. Comparative subcellular localization found 11 proteins as extracellular and cell wall proteins. Only 2 proteins were predicted to be virulent (MgpC protein and ab uncharacterized extracellular protein). The MgpC protein has a molecular weight > 110 kDa, and is thus excluded from the list. No T-cell epitopes were predicted for the uncharacterized extracellular protein, which was consequently also excluded from the list.

4. Discussion

Progress in sequencing technologies has advanced our understanding regarding the genomic information of microbial pathogens and opens up new avenues for targeting them [20];[21]. One route to combat bacterial pathogens is the development of an effective vaccine. The traditional vaccinology approach introduced by Pasteur has been successfully applied to many pathogens, and proved fruitful in providing protection against them [95]. However, classical vaccinology approaches face many obstacles during the development pipeline [22]. The shortcomings of traditional vaccinology are largely circumvented by integrating *in silico* filters, which compute these parameters with accuracy and in less time [22, 23, 24, 25]. Computational vaccinology has attained great focus in recent years due to it's user-friendly and straight forward nature [27]. The lack of a licensed vaccine against *P. stuartii* demonstrates the need for the development of an effective peptide-based vaccine. Peptide-based vaccines offer many potential advantages compared to traditional whole genome based vaccines [93]. They allow the host immune system to focus solely on a relevant antigenic peptide, thus avoiding the generation of non-protective responses, immune evasion, and autoimmune responses [58]; [93]. Similarly, peptide based vaccines are safer and could induce allergen specific tolerance. As peptide-based vaccines are more favorable and specific in their actions [58], we focused on the identification of antigenic

epitopes in the proteome of *P. stuartii*. One of the essential properties of vaccine proteins is virulence. Virulent proteins tend to initiate infection pathways more efficiently compared to non-virulent proteins [22, 23, 24, 25]. Before screening for virulent proteins, it was critical to discard those proteins, which were paralogous, host homologous and non-essential, residing in the cytoplasm or periplasm [25]. Removing host homologous proteins was essential as such proteins can mediate autoimmune responses due to cross-reactivity with the host proteins [22]. Target proteins must also be part of pathways essential for the survival of the pathogen. Blocking essential pathways of the bacteria is lethal for overall survival of the pathogen [25]; [35]. Targeting extracellular and outer membranous proteins was important, as they can be effective vaccine candidates due to their role pathogen virulence [22, 23, 24, 25]. Subsequent studies of the proteins revealed a set of thirteen virulent proteins: FimD1, FimD2, FimD3, FimD4, FimD5, FimD6, FimD7, FimD8, TolC, OmpF1, OmpF2, TC.OOP and MepM. All these 13 virulent proteins were human non-homologous, essential for pathogens survival and outer membrane localized. Further characterization of the targets based on their role in pathogen significant pathways revealed by KAAS reduced this subset to five proteins. These proteins are associated specifically with unique pathways of the pathogen (bacterial secretion system, two-component system, plant-pathogen interaction, β-lactam resistance, cationic antimicrobial peptide, and pili formation) and can be named as FimD1, FimD4, FimD6, FimD7, and FimD8 [49]. Screening for favorable physicochemical properties is also important for vaccine design during experimental analysis. The most critical parameter for a protein to act as vaccine target is their antigenic nature [22]. All the five proteins from the previous phase were found to be highly antigenic, and thus forwarded to subsequent analysis. Proteins containing large numbers of transmembrane helices are difficult to clone and express [22, 23, 24, 25], therefore, proteins with high number of transmembrane helices were not considered. All the five proteins were found to have one transmembrane helix. The molecular weight of all the targeted proteins was $\leq$ 110 kDa and considered due to their easy purification. The proteins were also found adhesive, which again could increase the virulence of *P. stuartii* by adhering and colonizing to host tissues [22, 23, 24, 25].

Mapping of B-cell derived T-cell epitopes for antigenic proteins is a critical step while designing vaccines [22, 23, 24, 25]. Only three epitopes: FVINDLYPT, IQWGNSHSA and YGNANVGYS from FimD4, FimD6 and FimD8 were considered, due to their higher binding affinity for MHC molecules, higher antigenic, virulence score, and low IC_{50} value. The proposed epitopes were all

identified as non-allergenic and conserved in strains of *P. stuartii*, important for designing a broad-spectrum vaccine [25]. These conserved regions evolve slowly and are subject to less variability, therefore, targeting such epitopes are reliable in the context of peptide-based vaccine design [24, 25]. The epitope sequence 'FVINDLYPT' (320-328) of FimD4 and 'IQWGNSHSA' (817-825) of FimD6 was conserved in 67%, thus can be used for designing strain-specific vaccines while epitope sequence 'YGNANVGYS' of FimD8 (677-685) was predicted to be conserved in all the 3 strains (100% conservation), a protein target for broad-spectrum vaccine design [25]. Furthermore, pepitope analysis of all the three epitopes revealed their surface exposed topology, which again favored them as suitable vaccine candidates. Molecular docking and complex analysis, showed the deep binding of the epitopes in the binding cavity of the receptor allele, making multiple hydrogen bonds and hydrophobic interactions with active site residues. The formation of multiple hydrogen bonds indicates the stability of the docked complex. Lastly, the pipeline was validated through a positive (*H.pylori*) and negative (*M. pneumoniae*) control. Studies on immune protective efficacy of vacA in small and large animal models revealed that this protein could induce local gastric T cell (Th1 and Th17) responses [22, 23]. In addition to vacA, the current pipeline predicted 4 other proteins (babA, fecA omp16, and sabA), which could serve as potential vaccine candidates against *H. pylori*. The *M. pneumoniae* study revealed no proteins that could serve as potential candidate. This was obvious as *M. pneumonia* lacks a cell wall and has very few cell surface proteins. The only two proteins, MgpC protein and uncharacterized extracellular protein, were found to have high molecular weight and did not bear any B-cell derived T-cell epitope, therefore, were not considered. Thus, we are confident in the predictions made by the designed pipeline. Despite of the many advantages of predicting antigenic peptides in bacterial proteomes, there remains several obstacles to effective vaccine development. These obstacles include: the low intrinsic immunogenicity of individual peptides, optimal delivery routes, the task of combining different antigenic peptides to properly engage cellular and humoral immunity, and poor population coverage of T-cell epitopes. Moreover, the prediction of B-cell epitopes are notoriously unreliable [25]; [93]. As there is a scarcity of antigenic epitopes against the pathogen, we believe that outcomes of this study will help experimental vaccinology in designing vaccine against this emerging bacterial pathogen and reducing the emergence and dissemination of its resistance strains.

Conclusions

Antimicrobial resistance by bacterial pathogens is a serious concern around the globe. Due to the slow pace of antibiotic development and the rapid pace of resistance acquisition, there is an urgent need for effective vaccines against these bacterial pathogens. In this context, we have identified novel potential vaccine candidates against *P. stuartii.* The proteins (FimD4, FimD6, and FimD8) are host non-homologous, essential, and part of *P. stuartii* exoproteome and secretome. Metabolic pathway mapping and the cellular interactome of the proteins indicated their pivotal role in the establishment of infection and its progression. B-cell derived T-cell epitopes were mapped successfully for the proteins, and found to be highly conserved and virulent. Furthermore, the antigenic epitopes demonstrated high affinity for the catalytic cavity of DRB1*0101, signifying that the epitopes could generate the desired response. These targets could be useful in designing broad-spectrum and strain-specific vaccines against virulent and multidrug-resistant clones of the pathogen.

Supplementary files

S-Fig. 1. Secondary structure of FimD4 protein.

S-Fig. 2. Secondary structure of FimD6 protein.

S-Fig. 3. Secondary structure of FimD8 protein.

S-Table 1: Pathway analysis of virulent proteins.

S-Table 2: Protein-protein interactions of five proteins associated with pathogen metabolic pathways.

S-Table 3: Structure evaluation of three targeted vaccine proteins.

S-Table 4: Five potential vaccine proteins against *H. pylori*.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

Authors are highly grateful to the Pakistan-United States Science and Technology Cooperation Program for granting the financial assistance

References

[1] G. Rahav, E. Pinco, F. Silbaq, H. Bercovier, Molecular epidemiology of catheter-associated bacteriuria in nursing home patients., J. Clin. Microbiol. 32 (1994) 1031–1034.

[2] J.W. Warren, *Providencia stuartii*: a common cause of antibiotic-resistant bacteriuria in patients with long-term indwelling catheters, Rev. Infect. Dis. 8 (1986) 61–67.

[3] T.D. Woods, C. Watanakunakorn, Bacteremia due to *Providencia stuartii*: review of 49 episodes., South. Med. J. 89 (1996) 221–224.

[4] M. Yoh, J. Matsuyama, M. Ohnishi, K. Takagi, H. Miyagi, K. Mori, K.-S. Park, T. Ono, T. Honda, Importance of Providencia species as a major cause of travellers' diarrhoea, J. Med. Microbiol. 54 (2005) 1077–1082.

[5] O.R. Sipahi, S. Bardak-Ozcem, E. Ozgiray, S. Aydemir, T. Yurtseven, T. Yamazhan, M. Tasbakan, S. Ulusoy, others, Meningitis due to *Providencia stuartii*, J. Clin. Microbiol. 48 (2010) 4667.

[6] G.D. Overturf, J. Wilkins, R. Ressler, Emergence of resistance of *Providencia stuartii* to multiple antibiotics: speciation and biochemical characterization of Providencia, J. Infect. Dis. 129 (1974) 353–357.

[7] S. Ünverdi, H. Akay, M. Ceri, S. Inal, M. Altay, A.P. Demiroz, M. Duranay, Peritonitis due to Providencia stuartii, Perit. Dial. Int. 31 (2011) 216–217.

[8] P.R. Krake, N. Tandon, Infective endocarditis due to Providencia stuartii, South. Med. J. 97 (2004) 1022–1024.

[9] E.S. Crane, M. Shum, D.S. Chu, Case report: *Providencia stuartii* conjunctivitis, J. Ophthalmic Inflamm. Infect. 6 (2016) 29.

[10] R.R. Chamberland, E.M. TeKippe, C.-A.D. Burnham, D.J. Kennedy, Renal abscess caused by a Providencia stuartii isolate biochemically misidentified as Pasteurella, J. Clin. Microbiol. 51 (2013) 2775–2777.

[11] G. KEREN, D.L.J. TYRREL, Gram-negative septicemia caused by *Providencia stuartii*, Int. J. Pediatr. Nephrol. 8 (1987) 91–94.

[12] S.T. Llah, A.D. Salman Khan, A.J.A. Morrison, S. Jain, D. Hermanns, A case of purple urine bag syndrome in a spastic partial quadriplegic male, Cureus. 8 (2016).

[13] C.E. Armbruster, S.N. Smith, A. Yep, H.L.T. Mobley, Increased incidence of urolithiasis and bacteremia during Proteus mirabilis and *Providencia stuartii* coinfection due to synergistic induction of urease activity, J. Infect. Dis. 209 (2014) 1524–1532.

[14] I. Stock, B. Wiedemann, Natural antibiotic susceptibility of *Providencia stuartii, P. rettgeri, P. alcalifaciens and P. rustigianii* strains, J. Med. Microbiol. 47 (1998) 629–642.

[15] Y.-C. Mao, C.-L. Chang, Y.-C. Huang, L.-H. Su, C.-T. Lee, Laboratory investigation of a

suspected outbreak caused by *Providencia stuartii* with intermediate resistance to imipenem at a long-term care facility, J. Microbiol. Immunol. Infect. (2016).

[16] M. Tumbarello, R. Citton, T. Spanu, M. Sanguinetti, L. Romano, G. Fadda, R. Cauda, ESBL-producing multidrug-resistant *Providencia stuartii* infections in a university hospital, J. Antimicrob. Chemother. 53 (2004) 277–282.

[17] L. Galani, I. Galani, M. Souli, I. Karaiskos, E. Katsouda, E. Patrozou, F. Baziaka, C. Paskalis, H. Giamarellou, Nosocomial dissemination of Providencia stuartii isolates producing extended-spectrum β-lactamases VEB-1 and SHV-5, metallo-β-lactamase VIM-1, and RNA methylase RmtB, J. Glob. Antimicrob. Resist. 1 (2013) 115–116.

[18] K. Hayakawa, D. Marchaim, G.W. Divine, J.M. Pogue, S. Kumar, P. Lephart, K. Risko, J.D. Sobel, K.S. Kaye, Growing prevalence of *Providencia stuartii* associated with the increased usage of colistin at a tertiary health care center, Int. J. Infect. Dis. 16 (2012) e646--e648.

[19] M. Bhattacharya, A. Parakh, M. Narang, others, Tigecycline, J. Postgrad. Med. 55 (2009) 65.

[20] S. Bambini, R. Rappuoli, The use of genomics in microbial vaccine development, Drug Discov. Today. 14 (2009) 252–260.

[21] R. Rappuoli, Reverse vaccinology, a genome-based approach to vaccine development, Vaccine. 19 (2001) 2688–2691.

[22] A. Naz, F.M. Awan, A. Obaid, S.A. Muhammad, R.Z. Paracha, J. Ahmad, A. Ali, Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: a reverse vaccinology based approach, Infect. Genet. Evol. 32 (2015) 280–291.

[23] A. Hassan, A. Naz, A. Obaid, R.Z. Paracha, K. Naz, F.M. Awan, S.A. Muhmmad, H.A. Janjua, J. Ahmad, A. Ali, Pangenome and immuno-proteomics analysis of *Acinetobacter baumannii* strains revealed the core peptide vaccine targets, BMC Genomics. 17 (2016) 732.

[24] D. Barh, N. Barve, K. Gupta, S. Chandra, N. Jain, S. Tiwari, N. Leon-Sicairos, A. Canizalez-Roman, A.R. dos Santos, S.S. Hassan, others, Exoproteome and secretome derived broad spectrum novel drug and vaccine candidates in *Vibrio cholerae* targeted by Piper betel derived compounds, PLoS One. 8 (2013) e52773.

[25] S. Baseer, S. Ahmad, K.E. Ranaghan, S.S. Azam, Towards a peptide-based vaccine against *Shigella sonne*i: A subtractive reverse vaccinology based approach, Biologicals. (2017).

[26] M.M. Giuliani, J. Adu-Bobie, M. Comanducci, B. Aricò, S. Savino, L. Santini, B. Brunelli, S. Bambini, A. Biolchi, B. Capecchi, others, A universal vaccine for serogroup B meningococcus, Proc. Natl. Acad. Sci. 103 (2006) 10834–10839.

[27] A. Sette, R. Rappuoli, Reverse vaccinology: developing vaccines in the era of genomics, Immunity. 33 (2010), 530-541.

[28] D. Maione, I. Margarit, C.D. Rinaudo, V. Masignani, M.Mora, M. Scarselli, H.Tettelin, C.Brettoni, E.T. Iacobini, R. Rosini, N. D'agostino, Identification of a universal Group B streptococcus vaccine by multiple genome screen, Science. 309 (2005) 148-150.

[29] C. Thorpe, L. Edwards, R. Snelgrove, O. Finco, A.Rae, G. Grandi, R. Guilio, T. Hussell, Discovery of a vaccine antigen that protects mice from *Chlamydia pneumoniae* infection, Vaccine. 25 (2007), 2252–2260.

[30] R. Singh, N. Garg, G. Shukla, N. Capalash, P. Sharma, Immunoprotective efficacy of *Acinetobacter baumannii* outer membrane protein, FilF, predicted in silico as a potential vaccine candidate, Front Microbiol. 7 (2016) . 158.

[31] S. Vivona, F. Bernante, F. Filippini, NERVE: new enhanced reverse vaccinology environment, BMC Biotechnol. 6 (2006) 35.

[32] R.J. Clifford, J. Hang, M.C. Riley, F. Onmus-Leone, R.A. Kuschner, E.P. Lesho, P.E. Waterman, Complete genome sequence of *Providencia stuartii* clinical isolate MRSN 2154, J. Bacteriol. 194 (2012) 3736–3737.

[33] U. Consortium, others, UniProt: a hub for protein information, Nucleic Acids Res. (2014) gku989.

[34] Y. Huang, B. Niu, Y. Gao, L. Fu, W. Li, CD-HIT Suite: a web server for clustering and comparing biological sequences, Bioinformatics. 26 (2010) 680–682.

[35] G. Sanober, S. Ahmad, S.S. Azam, Identification of plausible drug targets by investigating the druggable genome of MDR *Staphylococcus epidermidis*, Gene Reports. 7 (2017) 147-153.

[36] M. Sémon, K.H. Wolfe, Consequences of genome duplication, Curr. Opin. Genet. Dev. 17 (2007) 505–512.

[37] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res. 35 (2007) D61--D65.

[38] R. Zhang, H.-Y. Ou, C.-T. Zhang, DEG: a database of essential genes, Nucleic Acids Res. 32 (2004) D271--D272.

[39] B. Shanmugham, A. Pan, Identification and characterization of potential therapeutic candidates in emerging human pathogen *Mycobacterium abscessus*: a novel hierarchical in silico approach, PLoS One. 8 (2013) e59126.

[40] Y.Y. Nancy, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S.C. Sahinalp, M. Ester, L.J. Foster, others, PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, Bioinformatics. 26 (2010) 1608–1615.

[41] C.-S. Yu, Y.-C. Chen, C.-H. Lu, J.-K. Hwang, Prediction of protein subcellular localization,

735 Proteins Struct. Funct. Bioinforma. 64 (2006) 643–651.

736 [42] C.-S. Yu, C.-W. Cheng, W.-C. Su, K.-C. Chang, S.-W. Huang, J.-K. Hwang, C.-H. Lu,
737 CELLO2GO: a web server for protein subcellular localization prediction with functional
738 gene ontology annotation, PLoS One. 9 (2014) e99368.

739 [43] M. Mora, C. Donati, D. Medini, A. Covacci, R. Rappuoli, Microbial genomes and vaccine
740 design: refinements to the classical reverse vaccinology approach, Curr. Opin. Microbiol. 9
741 (2006) 532–536.

742 [44] L. Chen, D. Zheng, B. Liu, J. Yang, Q. Jin, VFDB 2016: hierarchical and refined dataset
743 for big data analysis—10 years on, Nucleic Acids Res. 44 (2016) D694--D697.

744 [45] C.E. Zhou, J. Smith, M. Lam, A. Zemla, M.D. Dyer, T. Slezak, MvirDB—a microbial
745 database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence
746 applications, Nucleic Acids Res. 35 (2007) D391--D394.

747 [46] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: an automatic genome
748 annotation and pathway reconstruction server, Nucleic Acids Res. 35 (2007) W182--W185.

749 [47] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, KEGG: Kyoto
750 encyclopedia of genes and genomes, Nucleic Acids Res. 27 (1999) 29–34.

751 [48] S.S. Azam, A. Shamim, An insight into the exploration of druggable genome of
752 *Streptococcus gordonii* for the identification of novel therapeutic candidates, Genomics.
753 104 (2014) 203–214.

754 [49] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M.
755 Simonovic, A. Roth, A. Santos, K.P. Tsafou, others, STRING v10: protein--protein
756 interaction networks, integrated over the tree of life, Nucleic Acids Res. (2014) gku1003.

757 [50] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens,
758 tumour antigens and subunit vaccines, BMC Bioinformatics. 8 (2007) 4.

759 [51] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch,
760 Protein identification and analysis tools on the ExPASy server, Springer, 2005.

761 [52] A. Krogh, B. Larsson, G. Von Heijne, E.L.L. Sonnhammer, Predicting transmembrane
762 protein topology with a hidden Markov model: application to complete genomes, J. Mol.
763 Biol. 305 (2001) 567–580.

764 [53] G.E. Tusnady, I. Simon, The HMMTOP transmembrane topology prediction server,
765 Bioinformatics. 17 (2001) 849–850.

766 [54] Y. He, Z. Xiang, H.L.T. Mobley, Vaxign: the first web-based vaccine design program for
767 reverse vaccinology and applications for vaccine development, Biomed Res. Int. 2010
768 (2010).

769 [55] J. Pizarro-Cerdá, P. Cossart, Bacterial adhesion and entry into host cells, Cell. 124 (2006)
770 715–727.

[56] Y. El-Manzalawy, D. Dobbs, V. Honovar, BCPREDS: B-cell epitope prediction server, Artif. Intell. Res. Lab. Dep. Comput. Sci. Iowa State Univ. Sci. Technol. Available Http//ailab. Cs. Iastate. Edu/bcpreds/. Accessed August. (2012).

[57] H. Singh, G.P.S. Raghava, ProPred1: prediction of promiscuous MHC Class-I binding sites, Bioinformatics. 19 (2003) 1009–1014.

[58] H. Singh, G.P.S. Raghava, ProPred: prediction of HLA-DR binding sites, Bioinformatics. 17 (2001) 1236–1237.

[59] N. Noureen, M. Tariq, A. Farooq, A. Arif, H. Bokhari, In silico identification of receptor specific epitopes as potential vaccine candidates from *Vibrio cholerae* strains, Gene Reports. 4 (2016) 222–232.

[60] P. Guan, C.K. Hattotuwagama, I.A. Doytchinova, D.R. Flower, MHCPred 2.0, Appl. Bioinformatics. 5 (2006) 55–61.

[61] A. Garg, D. Gupta, VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens, BMC Bioinformatics. 9 (2008) 62.

[62] A. Patja, S. Mäkinen-Kiljunen, I. Davidkin, M. Paunio, H. Peltola, Allergic reactions to measles-mumps-rubella vaccination, Pediatrics. 107 (2001) e27--e27.

[63] H.X. Dang, C.B. Lawrence, Allerdictor: fast allergen prediction using text classification techniques, Bioinformatics. 30 (2014) 1120–1128.

[64] B. Knudsen, T. Knudsen, M.Flensborg, H. Sandmann, M. Heltzen, A. Andersen, M. Dickenson, J. Bardram, P.J. Steffensen, S. Monsted, CLC Sequence Viewer, A/S Cb, version, 6 (2011).

[65] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.

[66] K. Lagesen, P. Hallin, E.A. Rødland, H.-H. Stærfeldt, T. Rognes, D.W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes, Nucleic Acids Res. 35 (2007) 3100–3108.

[67] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406-425.

[68] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.

[69] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J.E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (2015) 845–858.

[70] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, SWISS-MODEL: an automated protein homology-modeling server, Nucleic Acids Res. 31 (2003) 3381–3385.

[71] L.J. McGuffin, J.D. Atkins, B.R. Salehe, A.N. Shuid, D.B. Roche, IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences, Nucleic

Acids Res. 43 (2015) W169--W173.

[72] U. Pieper, B.M. Webb, G.Q. Dong, D. Schneidman-Duhovny, H. Fan, S.J. Kim, N. Khuri, Y.G. Spill, P. Weinkam, M. Hammel, others, ModBase, a database of annotated comparative protein structure models and associated resources, Nucleic Acids Res. 42 (2014) D336--D346.

[73] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinformatics. 9 (2008) 40.

[74] R.A. Laskowski, E.G. Hutchinson, A.D. Michie, A.C. Wallace, M.L. Jones, J.M. Thornton, PDBsum: a Web-based database of summaries and analyses of all PDB structures, Trends Biochem. Sci. 22 (1997) 488–490.

[75] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007) W407--W410.

[76] C. Colovos, T.O. Yeates, ERRAT: an empirical atom-based method for validating protein structures, Protein Sci. 2 (1993) 1511–1519.

[77] D. Eisenberg, R. Lüthy, J.U. Bowie, VERIFY3D: Assessment of protein models with three-dimensional profiles, Methods Enzymol. 277 (1997) 396–404.

[78] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.

[79] I. Mayrose, O. Penn, E. Erez, , N.D.Rubinstein, T. Shlomi, N.T. Freund, E.M. Bublil, E. Ruppin, R. Sharan, J.M. Gershoni, E.Martz, Pepitope: epitope mapping from affinity-selected peptides, Bioinformatics. 23 (2007.) 3244-3246.

[80] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem., 31 (2010), pp. 455-461.

[81] H. Lee, L. Heo, M.S. Lee, C. Seok, GalaxyPepDock: a protein--peptide docking tool based on interaction similarity and energy optimization, Nucleic Acids Res. (2015) gkv495.

[82] R.A. Laskowski, M.B. Swindells, LigPlot+: multiple ligand--protein interaction diagrams for drug discovery, (2011).

[83] E.T. Saulino, E. Bullitt, S.J. Hultgren, Snapshots of usher-mediated protein secretion and ordered pilus assembly, Proc. Natl. Acad. Sci. 97 (2000) 9240–9245.

[84] V. Koronakis, J. Eswaran, C. Hughes, Structure and function of TolC: the bacterial exit duct for proteins and drugs, Annu. Rev. Biochem. 73 (2004) 467–489.

[85] M. Fernández-Mora, J.L. Puente, E. Calva, OmpR and LeuO positively regulate the Salmonella enterica serovar Typhi ompS2 porin gene, J. Bacteriol. 186 (2004) 2909–2920.

844 [86] A. Ropy, J.A. Ayala, The Effect on Peptidoglycan Composition of Uncharacterized Pae-
845 Ampc Mutants Probes Its Functionality as DD-Peptidase, Int. J. Microbiol. Res. ISSN.
846 (2015) 975–5276.

847 [87] J.A. Gaddy, A.P. Tomaras, L.A. Actis, The *Acinetobacter baumannii* 19606 OmpA protein
848 plays a role in biofilm formation on abiotic surfaces and in the interaction of this pathogen
849 with eukaryotic cells, Infect. Immun. 77 (2009) 3150–3160.

850 [88] E.R. Green, J. Mecsas, Bacterial Secretion Systems--An overview, Microbiol. Spectr. 4
851 (2016).

852 [89] S. Wang, Bacterial two-component systems: structures and signaling mechanisms, INTECH
853 Open Access Publisher, 2012.

854 [90] R. Lakshmi, K.S. Nusrin, G.S. Ann, K.S. Sreelakshmi, Role of beta lactamases in antibiotic
855 resistance: A review, Int. Res. J. Pharm. 5 (2014) 37–40.

856 [91] M.E. Reardon-Robinson, H. Ton-That, Assembly and function of Corynebacterium
857 diphtheriae pili, in: *Corynebacterium diphtheriae* Relat. Toxigenic Species, Springer, 2014:
858 pp. 123–141.

859 [92] N. Vaishnav, A. Gupta, S. Paul, G.J. John, Overview of computational vaccinology: vaccine
860 development through information technology, J. Appl. Genet. 56 (2015) 381–391.

861 [93] W. Li, M.D. Joshi, S. Singhania, K.H. Ramsey, A.K. Murthy, Peptide vaccine: progress and
862 challenges, Vaccines. 2 (2014) 515-536.

863 [94] K.G. Frey, K.A. Bishop-Lilly, H.E. Daligault, K.W. Davenport, D.C. Bruce, P.S. Chain,
864 S.R. Coyne, O. Chertkov, T. Freitas, J. Jaissle, others, Full-genome assembly of reference
865 strain *Providencia stuartii* ATCC 33672, Genome Announc. 2 (2014) e01082--14.

866 [95] D. Serruto, R. Rappuoli, Post-genomic vaccine development, FEBS Lett. 580 (2006) 2985–
867 2992.